

UNITED STATES PATENT APPLICATION FOR:

**METHOD AND APPARATUS FOR BANDWIDTH
EFFICIENT DISTRIBUTION OF CONTENT**

INVENTORS:

MARCEL-CATALIN ROSU
DEBANJAN SAHA
SAMBIT SAHU
ANEES A. SHAIKH

ATTORNEY DOCKET NUMBER: YOR920030508US1

CERTIFICATION OF MAILING UNDER 37 C.F.R. 1.10

I hereby certify that this New Application and the documents referred to as enclosed therein are being deposited with the United States Postal Service on October 31, 2003, in an envelope marked as "Express Mail United States Postal Service", Mailing Label No. EV 413180810 US, addressed to: Mail Stop Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.


Signature

Linda DeNardi
Name

October 31, 2003
Date of signature

MOSER, PATTERSON & SHERIDAN, LLP
595 Shrewsbury Avenue
Shrewsbury, New Jersey 07702
(732)530-9404

METHOD AND APPARATUS FOR BANDWIDTH EFFICIENT DISTRIBUTION OF CONTENT

BACKGROUND

[0001] The present invention relates generally to computer systems and computer networks, and relates more particularly to content distribution over computer networks. Specifically, the present invention relates to a method and apparatus for bandwidth efficient delivery of content over a computer network.

[0002] A number of internet applications work in a client-server mode wherein servers located in data centers distribute the same content or data (*i.e.*, “packets”) to a number of clients. Examples of such applications include audio/video streaming, software distribution and multiparty online gaming, among others. In most of these applications, although the servers send the same data to each client in a set of clients, the data is sent to one client at a time using a unicast distribution system.

[0003] Figure 1 is a schematic diagram illustrating one embodiment of a unicast distribution system. Such a method is also referred to as point-to-point communication. As illustrated, the server sends a plurality of copies of the same data to a plurality of clients using a unicast forwarding mechanism. The packets are typically distributed on a one-to-one basis; that is, an individual copy of the same information is produced and sent to each client (*e.g.*, a sender with ten receivers would send ten packets along ten routes). Consequently, a major cost associated with the distribution of content from server to client(s) is that of bandwidth. When the number of clients is particularly large, *e.g.*, in the case of multiparty online games, bandwidth costs may dominate the total cost of hosting the application.

[0004] One conventional way to reduce the usage and cost of bandwidth is to use a point-to-multipoint multicast distribution system. In a point-to-multipoint multicast distribution system, the receivers form a distribution tree wherein the sender is the “root” of the tree and the receivers are the “leaves”. Routers form intermediate nodes in the tree and are adapted to forward data items downstream, *i.e.*, towards the receivers. Thus, the

YOR920030508US1

sender sends a single copy of the data item, addressed to a group of receivers, and as the data is routed through the multicast distribution tree it is copied by the intermediate routers and sent over diverging branches of the distribution tree. Although such systems can potentially reduce the usage and cost of bandwidth, they are hampered by a lack of scalability and interoperability. Furthermore, such multicast systems typically require an overlay infrastructure, and the cost of deployment and maintenance for multicast overlays can be quite costly. Other solutions, such as peer-to-peer systems, must run slow and complex protocols to create multicast overlays and can not be easily reconfigured on a packet-by-packet basis. This creates difficulties when the multicast group membership changes or when system hardware (*e.g.*, intermediate receivers or routers) fail.

[0005] Thus, there is a need for a method and apparatus for a simplified, flexible, and bandwidth efficient system for point-to-multipoint distribution of content over a computer network.

SUMMARY OF THE INVENTION

[0006] In one embodiment, the present invention is a method and an apparatus for bandwidth efficient distribution of content over a computer network. In one embodiment, a server cooperates with clients to reduce the usage of server bandwidth, thereby reducing bandwidth-associated distribution costs. The server controls the way that content is distributed within the system, and the clients/receivers may act as intermediate forwarders of data packets. In one embodiment, this is accomplished by sending data items to a first group of receivers on a packet-by-packet basis using unicast communication. The first group of receivers then sends the packets on to a second set of receivers, and the process continues until the packet is forwarded to all intended receivers. Namely, each data packet contains a point-to-multipoint distribution sub-tree. Under this novel distribution scheme, different packets reach different receivers via different intermediate receivers, thereby reducing the usage and cost of bandwidth.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] So that the manner in which the above recited embodiments of the invention are attained and can be understood in detail, a more particular description of the invention, briefly summarized above, may be had by reference to the embodiments thereof which are illustrated in the appended drawings. It is to be noted, however, that the appended drawings illustrate only typical embodiments of this invention and are therefore not to be considered limiting of its scope, for the invention may admit to other equally effective embodiments.

[0008] Figure 1 is a schematic illustration of a conventional unicast distribution system;

[0009] Figure 2 is a schematic illustration of one embodiment of a point-to-multipoint distribution system according to the present invention;

[0010] Figure 3 is a flow diagram illustrating one method of distributing content using the system illustrated in Figure 2; and

[0011] Figure 4 is a block diagram of a system for executing the method in accordance with embodiments of the present invention.

[0012] To facilitate understanding, identical reference numerals have been used, where possible, to designate identical elements that are common to the figures.

DETAILED DESCRIPTION

[0013] The present invention is a method and an apparatus for bandwidth efficient distribution of content over a computer network. Figure 3 is a flow diagram illustrating one embodiment of a method 300 of distributing content. In step 302, a sender (*e.g.*, a server) encodes a first data packet (containing data to be distributed) with a multicast tree structure that specifies a list of receivers (or levels/groups of receivers) to which the first packet should be sent and a path along which the first packet should be sent. In step 304, the first packet is sent to a first group of receivers, for example, using a conventional unicast distribution system as described herein. In step 306, each receiver

in the first group of receivers, having received the first packet, makes a specified number of copies of the first packet, and in step 308 each receiver in the first group sends its packet copies to a second group of receivers as specified by the tree structure with which the first packet is encoded. In step 310, each receiver in the second group of receivers copies the first packet, and these copies made by the second group of receivers are forwarded to a subsequent group of receivers, for example using a unicast distribution system. Thus, steps 306 and 308 are essentially repeated $N-1$ times, wherein N is the number of groups of receivers to which the first packet must be forwarded. The multicast distribution tree structure will specify the paths (*i.e.*, the intermediate receivers) by which the first packet arrives at each receiver in the tree. Once the first packet has reached all receivers as defined by the multicast distribution tree or a portion thereof, the method 300 is repeated for subsequent data packets.

[0014] Figure 2 is a schematic illustration of one embodiment of a point-to-multipoint distribution tree structure 200 for carrying out the method 300 illustrated in Figure 3. In the embodiment illustrated, the system 200 is a two-level system that comprises a server 202, a first group of receivers 204a – 204c (hereinafter referred to as “receivers 204” and showing a “fan out” of three receivers), a second group of receivers 206a – 206c (hereinafter referred to as “receivers 206” and showing a “fan out” of three receivers), a third group of receivers 208a – 208c (hereinafter referred to as “receivers 208” and showing a “fan out” of three receivers), and a fourth group of receivers 210a – 210c (hereinafter referred to as “receivers 210” and showing a “fan out” of three receivers). Although four sets of receivers 204, 206, 208 and 210 are illustrated, those skilled in the art will appreciate that an infinite number of receivers and groups of receivers can be used to advantage with the present invention, and the grouping of receivers will depend on parameters determined by the server as described further herein.

[0015] In one embodiment, the sender (*e.g.*, server) 202 sends a packet p_1 intended for the receivers 204, 206, 208 and 210 (step 304). The packet p_1 is sent by the server 202 to the first group of receivers 204 using a unicast communication system as described herein. The packet p_1 contains data and the multicast distribution tree, or at

YOR920030508US1

least the relevant portion of the multicast distribution tree that defines a list of receivers to which the first group of receivers 204 should forward the packet p_1 (step 302). Each receiver 204 in the first group of receivers 204 makes the requisite number of copies of the packet p_1 (step 306) and sends a copy of the packet p_1 to each subsequent receiver 206, 208 or 210 in its distribution sequence or structure (step 308). Thus, receiver 204a forwards the packet p_1 to the second group of receivers 206, receiver 204b forwards the packet p_1 to the third group of receivers 208, and the receiver 204c forwards the packet p_1 to the fourth group of receivers 210. Forwarding from the receivers 204 to the second, third or fourth groups of receivers 206, 208 or 210 is accomplished using a unicast communication system.

[0016] In one embodiment, the server 202 determines and controls the distribution process for the packet p_1 , *e.g.*, the server 202 can alter the multicast distribution tree structure for each packet sent. The server 202 encodes the packet p_1 with the list of receivers 204, 206, 208 or 210 to which the packet p_1 is to be delivered, as well as with the order in which the packet p_1 is to be delivered to the individual receivers 204, 206, 208 or 210 (*i.e.*, which receivers forward the packet p_1 to which other receivers). Thus, the packet p_1 essentially contains a point-to-multipoint distribution sub-tree.

[0017] Although the embodiment illustrated in Figure 2 depicts a distribution system 200 having two levels of receivers (*i.e.*, receivers 204 and receivers 206/208/210) forwarding one packet p_1 of data, those skilled in the art will appreciate that the distribution system 200 may comprise an infinite number of receivers (*e.g.*, receiver group N in step 312 of Figure 3) for forwarding an infinite number of packets. The actual number of receivers (“fan out”) and levels into which the receivers are grouped (“depth”) will depend on application requirements or metrics including latency, reliability, incentive to conserve bandwidth, and available bandwidth at each receiver, among others. For example, for applications requiring low latency, small depth and fan out may be desired. If an application values low bandwidth costs over latency, deep trees with large fan out may be better suited to handle these requirements.

[0018] Thus, the number of permutations for packet delivery routes is a function of the number of receivers in the distribution tree. That is, using the distribution method illustrated in Figures 2 and 3, different packets may be sent to different receivers via different intermediate receivers. The server 202 has substantially complete control over the distribution process, and in one embodiment, the server 202 can change the distribution tree on a packet-by-packet basis. Consequently, even if receivers are spontaneously added to or dropped from the distribution tree, the variable structure of the tree can substantially reduce failures in packet delivery by simply re-routing packets to available receivers. For example, if receiver 204c is dropped from the system 200 illustrated in Figure 2, the receiver 204b can deliver the packet p_1 to the fourth group of receivers 210, as illustrated by the dotted lines connecting receiver 204b to the receivers 210. In one embodiment, packets are further encoded with forward error correction coding to further reduce packet loss. Therefore, each receiver need only receive one subset of packets in order to reconstruct the content. Furthermore, the sender 202 can adjust the structure of the multicast distribution tree to address different metrics, including, but not limited to, cost, delay and bandwidth constraints.

[0019] Figure 4 is a high level block diagram of the present invention implemented using a general purpose computing device 400. In one embodiment, a general purpose computing device 400 comprises a processor 402, a memory 404 and various input/output (I/O) devices 406 such as a display, a keyboard, a mouse, a modem, and the like. In one embodiment, at least one I/O device is a storage device (e.g., a disk drive, an optical disk drive, a floppy disk drive). In one embodiment, the method for content distribution described herein with reference to the preceding Figures can be implemented as a software application that is retrieved from the storage device 406 and is loaded onto the memory 404. The processor 402 can then execute the method in the memory in accordance with embodiments of the present invention. As such, it is contemplated that some and/or all of the steps of the methods and the data structure discussed above can be stored on a computer readable medium.

[0020] Thus, the present invention represents a significant advancement in the field of content distribution over computer networks. A point-to-multipoint system is provided that consumes significantly less bandwidth at a given time than conventional point-to-point systems. The server controls the destinations and routes for each packet that the server sends, and destinations and routes may be altered on a packet-by-packet basis to reduce packet loss. The system is simple and does not require deployment or maintenance of multicast overlays, which can be costly. The speed and simplicity of the system make the system robust against changes in multicast group membership or failure of individual receivers.

[0021] While foregoing is directed to the preferred embodiment of the present invention, other and further embodiments of the invention may be devised without departing from the basic scope thereof, and the scope thereof is determined by the claims that follow.